

BMJ Open In primary care, is measuring free-thyroxine plus thyroid-stimulating hormone to detect hypopituitarism cost-effective? A cost utility analysis using Markov chain models

Brian Shine,¹ Tim James, Amanda Adler²

To cite: Shine B, James T, Adler A. In primary care, is measuring free-thyroxine plus thyroid-stimulating hormone to detect hypopituitarism cost-effective? A cost utility analysis using Markov chain models. *BMJ Open* 2019;9:e029369. doi:10.1136/bmjopen-2019-029369

► Prepublication history for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-029369>).

Received 23 January 2019
Revised 14 June 2019
Accepted 08 July 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Clinical Biochemistry, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

²Institute of Metabolic Sciences, Addenbrooke's Hospital, Cambridge, UK

Correspondence to

Dr Brian Shine;
brianshine@mac.com

ABSTRACT

Objective We examined whether it is cost-effective to measure free thyroxine (FT4) in addition to thyrotropin (thyroid-stimulating hormone (TSH)) on all requests for thyroid function tests from primary care on adult patients.

Background Hypopituitarism occurs in about 4 people per 100 000 per year. Loss of thyrotropin (TSH) secretion may lead to secondary hypothyroidism with a low TSH and low FT4, and this pattern may help to diagnose hypopituitarism that might otherwise be missed.

Design Markov model simulation.

Primary outcome measure Incremental cost-effectiveness ratio (ICER), the ratio of cost in pounds to benefit in quality-adjusted life years of this strategy.

Results The ICER for this strategy was £71 437. Factors with a large influence on the ICER were the utilities of the treated hypopituitary state, the likelihood of going to the general practitioner (GP) and of the GP recognising a hypopituitary patient. The ICER would be below £20 000 at a cost to the user of an FT4 measurement of £0.61.

Conclusion With FT4 measurements at their present cost to the user, routine inclusion of FT4 in a thyroid hormone profile is not cost-effective.

INTRODUCTION

Hypothyroidism is common in the UK, but hypothyroidism secondary to hypopituitarism is not. Primary hypothyroidism (caused by disease of the thyroid itself) occurs in approximately 400 per 100 000 women per year and 60 per 100 000 men per year.¹ By contrast, only about 4 per 100 000 people per year develop hypothyroidism secondary to disease of the pituitary or hypothalamus.^{2,3}

In primary hypothyroidism, the concentrations of serum or plasma free thyroxine (FT4) and free triiodothyronine (FT3) are low reflecting the thyroid's inability to produce the adequate hormone, whereas (in response) the concentration of thyroid-stimulating hormone (TSH, thyrotropin) is high. In secondary hypothyroidism, for a given

Strengths and limitations of the study

- The first study to use a health-economic approach to determine the cost-effectiveness of including free thyroxine (FT4) in routine thyroid function tests for specimens from primary care.
- This method has general applicability for diagnostic tests and allows the effects of factors such as cost per test, sensitivity and specificity to be analysed.
- Because there are no randomised studies of either the primary intervention (measurement of FT4) or the effect of treatment of hypopituitarism, our analysis depends on observational data.

concentration of FT4, the TSH is inappropriately low, reflecting the pituitary's inability to produce adequate hormone.⁴

When general practitioners (GPs) in the UK who suspect hypothyroidism request 'thyroid function tests' (TFTs) in adults, many National Health Service (NHS) laboratories in the UK measure TSH alone, and add FT4 only when the TSH result is abnormal, that is, too low or too high or close to the lower or upper end of the reference interval. However, this strategy will fail to identify people with hypothyroidism caused by hypothalamic or pituitary disease, whose TSH values appear 'normal' when looked at in isolation, but are inappropriately low for the concentration of FT4. Identifying people with undetected pituitary disease is likely to improve both quality of life and length of life as people with untreated pituitary disease have higher mortality rates than patients with treated hypothalamic or pituitary disease, even when the TSH is normal and/or they do not have symptoms.⁵ Because pituitary disease may cause non-specific symptoms, such as tiredness, some clinicians have called

for measuring FT4 in addition to TSH when screening for thyroid dysfunction.

We have examined whether it is cost effective in the UK to screen for pituitary disease by measuring FT4 in addition to TSH in blood samples for TFTs from primary care. We model a cohort of patients using a cost-utility analysis over the lifetime of patients to compare each strategy to determine the costs, quality of life and length of life.

METHODS

Clinical background

Disorders of the hypothalamus, pituitary stalk or pituitary itself may cause deficiency of one or more pituitary hormones (hypopituitarism). Depending on the underlying disease, this may be accompanied by excess secretion of other pituitary hormones, such as prolactin, growth hormone (GH) or adrenocorticotropin (ACTH). ACTH and TSH secretion tend to be lost later than other hormones in hypopituitarism. Although patients with excess hormone secretion may have recognisable clinical features and premenopausal females may develop amenorrhoea, other groups may have only non-specific features, such as tiredness. It is important to diagnose pituitary disorders: the underlying cause may be treatable, for instance, by surgery or appropriate medications such as dopamine agonists or somatostatin analogues; treating hypopituitarism reduces the excess morbidity and mortality associated with this condition.

Laboratory tests for suspected pituitary disease include measuring follicle-stimulating hormone (FSH), luteinising hormone (LH) and prolactin, but assessing pituitary function for ACTH, TSH and GH secretion usually requires dynamic function tests, such as inducing hypoglycaemia with insulin. In TSH deficiency, the TSH may be low or undetectable. However, if some TSH secretion is retained, but the pituitary's response to low thyroid hormones is inadequate, there should be a mismatch between the TSH and FT4.⁴

Population

We developed a model using data from a population-based laboratory database for 2016 to calculate the age and sex distribution of adults (age ≥ 18 years) who have thyroid tests in Oxfordshire. The total population in the catchment is approximately 680 000; the population aged ≥ 18 years is 540 272 (272,717 women). In 2016, this reference laboratory performed 191 043 thyroid tests on specimens from primary care in this population.

To calculate the death rates given the mean age of the adult population (51 years) and stratified by sex, we used UK mortality tables. We estimated the incidence of hypopituitarism in the population to be 4.2 per 100 000 per year, with 50% of cases occurring in women.²

Patient and public involvement

There was no formal patient or public involvement in the development of the research question.

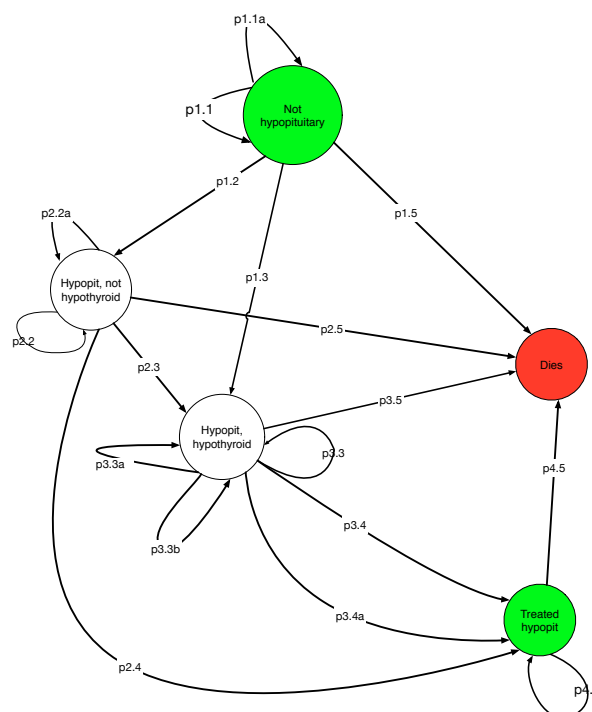


Figure 1 Markov chain model with five health states, showing possible transitions between them.

Model

We simulated the clinical question by developing a discrete Markov chain model with five health states (figure 1) through which we sent a hypothetical population of patients. The Markov chain describes a sequence of possible events and connects them with the probability, over time, of moving from one health state to the next. The health states are:

1. No pituitary disease.
2. Undiagnosed pituitary disease without hypothyroidism.
3. Undiagnosed pituitary disease with undiagnosed secondary hypothyroidism.
4. Treated pituitary disease with or without secondary hypothyroidism.
5. Death, an absorbing state.

We chose a cycle length of 28 days. At the end of each cycle, a (simulated) person may stay in his or her present state or move to a higher numbered state. We chose values to represent the probability of moving from state to state (see transition matrices below). In the first three states, the patient may visit his/her GP because of symptoms and the GP may request tests, including TFTs. We modelled two different options based on a laboratory's policy of whether it measures FT4 automatically or measures it only in response to a TSH near either end of the reference interval or outside the reference interval. The probabilities of the various events are summarised in table 1.

GPs request TFTs in as many as 15.7% of adults per year.⁶ This figure is similar to those for our own laboratory, so

Table 1 Probabilities and RRs used

Node	Label	Description	Probability (%) or RR	Source
1	p1.5	Well to death	From mortality tables	From mortality tables
1	p1.2	Become hypopituitary, not hypothyroid	0.0021/year	Ref 2
1	p1.3	Become hypopituitary, hypothyroid	0.0021/year	Ref 2
1	p1.1a	GP requests TFTs	16/year	Ref 6
2	p2.5	Relative risk of death with hypopituitarism	2.4	Ref 5
2	p2.3	Hypopituitary to hypothyroid	50/year	Assumed
2	p2.4	GP recognises hypopituitary	20	Assumed
2	p2.2	Does not visit GP	10	Assumed
3	(p3.5)	RR of death with with hypopit/hypothyroid	3.6	Assumed
3	p3.4.t4	Tests suggest hypopituitary (with T4)	64	Ref 3
3	p3.4.tsh	Tests suggest hypopituitary (no T4)	1	Ref 3
3	p3.4a	GP recognises hypopituitary	50	Assumed
3	p3.3a	Visit GP, no TFTs	5	Assumed
3	p3.3	Does not visit GP	10	Assumed
3	p3.4.t4f	False positive rate	0.3	Refs 3, 16
4	p4.5	RR of death on treatment	1.1	Ref 5

GP, general practitioner; RR, relative risk; TFT, thyroid function test.

we have assumed that GPs request 16 sets of TFTs per 100 adults per year.

Transition matrices

We defined two transition matrices, which represent the probabilities of progressing from one health state to another. The matrices reflect the different options: one matrix reflects testing TSH alone in specimens from primary care, the other includes FT4. The only difference between the two matrices is the difference in the probability of diagnosing central hypothyroidism (moving from the state of hypopituitarism with hypothyroidism to the state of diagnosed hypopituitarism).

Costs (table 2)

The current additional cost of adding a FT4 test to TSH in our laboratory is £4.00. In the reference laboratory, about 25% of specimens from primary care already have an FT4 added because of a low or high TSH.

A false-positive result occurs when a patient in state 1 (no pituitary disease) has a result that suggests pituitary disease. We assume that the proportion of false positive results is 0.3%. This figure is derived from the paper of Preiss and colleagues.³ We estimate that the tests used to exclude hypopituitarism in 2019 after a false-positive result cost £ 50.00. This figure is estimated from repeating the TFTs, and measuring GH, FSH, LH and prolactin, and cortisol on a specimen obtained at 9 a.m., based on the costs from our laboratory.

We estimate that the tests to confirm hypopituitarism cost approximately £1000. This includes the costs of visits

to the endocrine clinic (approximately £500, based on an initial visit costing £200, plus a follow-up visit costing £100, and dynamic tests in the endocrine unit £300, based on costs in our Trust), plus any further tests that are required, such as a short synacthen test and pituitary imaging. The cost of a year of pituitary replacement therapy is estimated to be £2000. The major component of this is the cost of GH replacement, which is currently about £20.00 per mg for biosimilars.⁷ The average GH dosage has been estimated to be 0.29 mg/day,⁸ giving an average cost for GH replacement of £2117 for those on GH. We assume that 75% of hypopituitary patients receive GH. Other costs, including other hormone replacements (hydrocortisone, thyroxine and testosterone/sex steroids), plus testing and other clinical management are assumed to be contained in the overall figure.

Mortality

In the baseline model, we use an average age at diagnosis of 46 years.⁹ We use the population mortality provided by the UK Office for National Statistics for 2013–2015.¹⁰ We use a lifetime horizon to age 100 to capture all costs and benefits.

In a meta-analysis, Pappachen and colleagues⁵ examined the effect of GH replacement in hypopituitarism, and estimated the standardised mortality rate (SMR) among patients with hypopituitarism comparing people not on GH and people on growth hormone replacement to the general population. The SMR for people not on growth hormone therapy compared with the general

population was 2.40 (95% CI: 1.46 to 3.34; 3.04; 95% CI: 1.98 to 4.10 in women vs 2.09; 95% CI: 1.24 to 2.94 in men), whereas the SMR for people on GH replacement was 1.15 (95% CI: 1.05 to 1.24; 1.57; 95% CI: 1.38 to 1.77 in women and 0.95; 95% CI: 0.85 to 1.06 in men). In the absence of information on covariates, we use the crude figures.

The Birmingham group⁹ did not find evidence that thyrotropin deficiency increased mortality over the baseline mortality, but, in this scenario, we assume that the mortality is twice as high in those with hypopituitarism and secondary hypothyroidism than in those with hypopituitarism without hypothyroidism.

Utilities

We used the utility values obtained by Koltowska-Haggstrom and colleagues using UK tariffs.¹¹ This study derived values for health-related quality of life from the general population of England and Wales by regressing a 25-item quality of life measure specific for GH deficiency in adults against the EuroQual 5-D.

With a value of 1 reflecting perfect health and value of 0 reflecting death, we estimate that the utility for the normal population (health state 1) is 0.90, for hypopituitarism (state 2) 0.67, for hypopituitarism with hypothyroidism (state 3) 0.62 and for patients on pituitary replacement therapy (state 4) 0.75.¹¹

We have assumed that a false-positive test does not significantly reduce the quality of life.

We assume a discount rate of 3.5% for both costs and benefits.

Running the model

Each simulation starts with a probability of 1 of being in state 1. At the end of each cycle, the probabilities of moving to another state are calculated. The cycles are repeated until the horizon is reached. The probability of being in each state is multiplied by the utility or the costs, with discounting. The sum of these gives the relevant quality-adjusted life years (QALYs) and the costs. The difference in costs divided by the difference in QALYs gives the incremental cost-effectiveness ratio (ICER).

For the probabilistic simulation, we repeat the process while randomly varying inputs as explained below.

By multiplying the costs of measuring FT4 by the probability of doing TFTs in states 1, 2 and 3, we calculated the extra cost incurred of adding FT4. In state 1, we calculated the cost of false-positive tests by multiplying three values: the probability of requesting a TFT, the probability of false positive results and the cost of follow-up tests. We assumed that one set of tests and their associated costs were used to confirm hypopituitarism (when moving from state 2 or 3 to state 4).

We estimated the ICER by calculating the difference in total costs (in £) between the groups with FT4 plus TSH compared with measuring TSH alone divided by the difference in QALYs (Δ costs/ Δ QALY).

Table 2 Costs used in the model

Cost source	Cost (£)	Source of the estimated cost
FT4	4.00	Current cost in our laboratory
False-positive result	50.00	Estimated by authors
Confirmatory tests	1000	Estimated using current costs in our Trust
Cost of GH replacement	2117	References 7, 8

FT4, free thyroxine; GH, growth hormone.

Sensitivity analysis

We examined the effect on the baseline ICER by varying the time horizon and the probabilities, rates and costs in tables 1 and 2 by 10%.

Probabilistic analysis

We conducted a probabilistic analysis by simulating variation that would occur randomly in the transition probabilities, utilities and costs. We used beta pert distributions for probabilities, utilities, costs and multipliers.¹² For age at which thyroid functions were done, we used random sampling from the age distribution of the adult population having TFTs in the reference laboratory. We ran 10 000 simulations.

RESULTS

The base case model gives a difference in QALYs of 0.000157 at a cost of £11.24, that is a cost per QALY of £71 437 (table 3).

The ICER is linearly related to the cost of an FT4 test, with an intercept of £10 781 per QALY gained and a slope of £15 164 per QALY gained per £1 increase in the cost of a test. This means that the total extra cost of an FT4 to the requestor would have to fall to £0.61 to make the ICER fall below £20 000 per QALY gained.

Sensitivity analysis

Deterministic analyses

A sensitivity analysis tests which inputs impact the results. The deterministic approach varies one variable at a time from a prespecified range, in this case plus or minus 10%. From table 4 and figure 2, the most influential variables are the utilities of the treated hypothyroid

Table 3 Base case ICER calculation

Cost (£)	QALYs	Δ Cost (£)	Δ QALYs	ICER
23.25	17.891419			
34.49	17.891576	11.24	0.000157	71 437

ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year.

Table 4 Exploration of effect of varying factors by ± 10

Factor	Difference (%)	Sign
Utility (treated hypopituitarism)	58.12	-
Utility (hypopituitarism)	37.90	+
p (GP diagnoses state 3)	37.51	+
p (develop hypopituitarism)	18.10	-
p (GP requests TFTs, general)	17.92	+
FT4 cost	16.98	+
Age at onset of hypopituitarism	15.51	-
SMR (central hypothyroidism)	14.76	-
p (go to GP, state 3)	13.87	+
p (FT4 abnormal, state 3)	12.17	-
p (GP requests TFTs, state 3)	11.73	-
SMR (hypopituitarism on treatment)	10.14	+
Discount rate	7.44	+
Utility decline (central hypopituitarism)	2.73	-
p (hypopit to hypothyroid)	2.33	-
Cost of pituitary replacement	2.01	+
p (go to GP, state 2)	1.78	+
p (GP diagnoses state 2)	1.78	+
FP rate (FT4)	0.94	+
Cost (FP)	0.94	+
p (TSH abnormal, state 3)	0.43	+
SMR (hypopituitarism)	0.35	+
p (central hypothyroidism)	0.00	0
Utility (baseline)	0.00	0

FP, false positive; FT4, free thyroxine; GP, general practitioner; SMR, standardised mortality rate; TFT, thyroid function test.

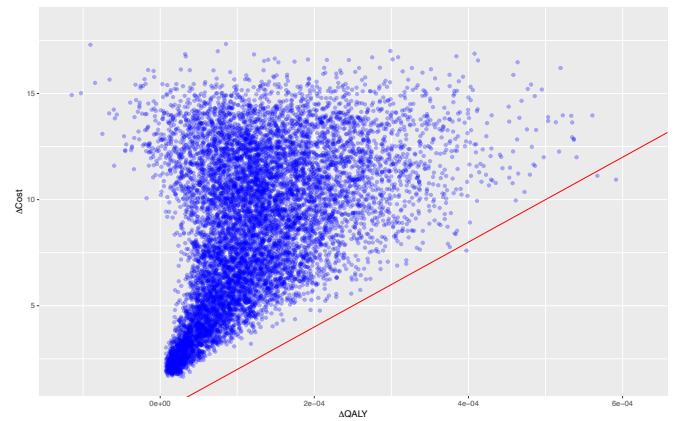


Figure 3 Probabilistic sensitivity analysis (with ICER of £20 000 per QALY gained shown as red line). ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year.

state, the hypopituitary state, the likelihood of going to the GP and of the GP recognising a hypopituitary patient.

Probabilistic analysis

A probabilistic analysis changes the values of the inputs to determine what happens to cost-effectiveness. The source of the values come from random samples from each of the distributions of the variables. The mean ICER was £122 417 per QALY gained (median, £73,242; IQR, £51 593–1 08 016).

The result of this analysis is shown graphically in figure 3, where the ICER is plotted on the cost-benefit plane, with a reference line with a slope of £20 000 per QALY gained. Only 0.9% of these calculated ICER values are <£20 000 per QALY gained.

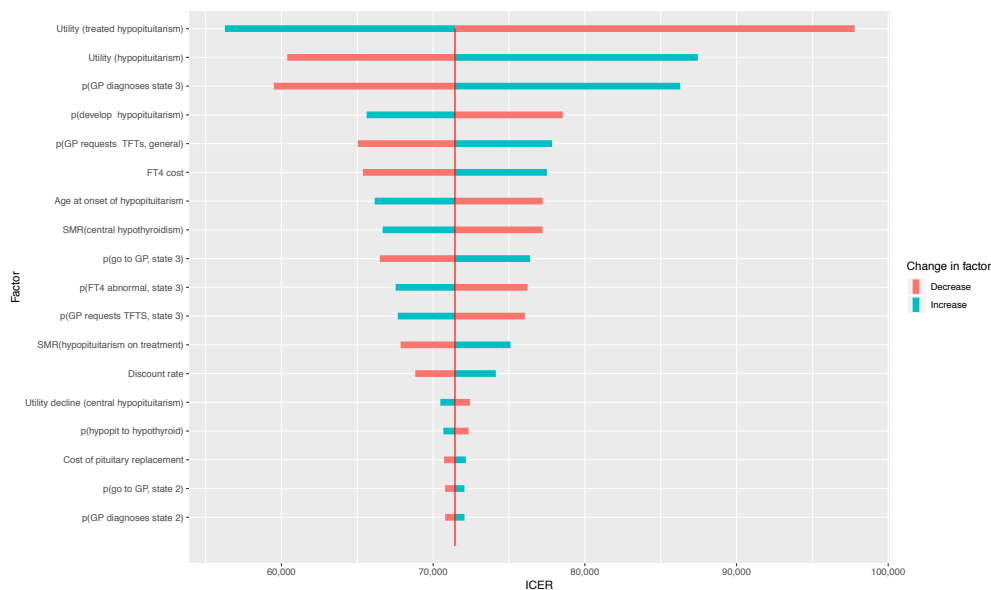


Figure 2 Tornado diagram, showing the effect on the ICER calculation of altering each factor by 10%. ICER, incremental cost-effectiveness ratio.

Budget impact

The budget impact estimates the total cost, rather than cost-effectiveness. If 15% of adults have a TFT per year, and a quarter of these would have an FT4 in any case, and the adult population of England is about 40 million, then this would imply that laboratories would perform an extra 4.5 million FT4 tests compared with a policy of just analysing TSH, costing about £19.1 million per year.

Because more cases of hypopituitarism would be diagnosed, the costs of treatment would also rise.

DISCUSSION

A survey of laboratory practice in the 1990s¹³ suggested that most laboratories in the UK use TSH as the front-line test for requests for thyroid function requests from primary care, adding FT4 and possibly FT3 if the TSH falls outside certain limits. However, this strategy does not identify people with hypopituitarism who have a 'normal' TSH that is inappropriately low for their (unmeasured) FT4. In addressing whether it would be reasonable from a clinical and cost standpoint to measure FT4 at the onset, we conclude that laboratories should not routinely test specimens for FT4 in specimens from patients requested by GPs.

Preiss and colleagues³ studied 73 650 requests for TFTs from primary care and identified 10 new cases of hypopituitarism. They concluded that measuring FT4 added to TSH is clinically effective, but did not estimate the costs or the benefits of this strategy.

Livingston and colleagues¹⁴ examined 26 106 consecutive TFTs and identified eight patients with hypopituitarism who would otherwise have been missed without an FT4 measurement. They concluded that the extra reagent cost of identifying each case was about £1451 based on an approximate estimate per FT4 test of £0.45. They took into account only marginal reagent costs to the laboratory and did not include non-laboratory costs such as imaging and the costs of false-positive tests (113 patients in this study, approximately 0.5% of tests). Because the study was cross-sectional, it estimates the prevalence rather than the incidence of hypopituitarism, and thus underestimates the costs of an ongoing policy of using FT4. Additionally, the true laboratory cost of the FT4 is greater than the marginal costs of FT4 reagent alone: it will include many additional elements including consumable costs, a proportion of the automation hardware costs and a range of laboratory staff costs, such as increased reagent management and data review.

Hypopituitarism is a comparatively rare condition. The main benefit of measuring FT4 routinely is to make the diagnosis earlier. During the delay incurred by not measuring FT4, the risk of death is relatively small, so the main benefit is in quality of life over this extra period. These factors mean that the overall health gains for the whole population by using FT4 as a screening test are small, and so the ICER is high. We, therefore, suggest that it does not reflect good value for money.

The cost-ineffectiveness of screening for thyroid disease secondary to pituitary disease depends most heavily on the estimates of health-related quality of life associated with treated and untreated disease. However, these utilities are uncertain. A systematic review,¹⁵ on which we base our estimates, summarised studies of quality of life in hypopituitarism. Many studies required mapping the results of assessments using instruments specific to pituitary disease to estimates of utility (notably, the EQ-5D). In one study,¹¹ the authors used a GH-specific quality of life instrument in patients and mapped this to EQ-5D from the answers to the two questionnaires in 921 healthy individuals in England and Wales. The derived utilities for the non-hypopituitary population were 0.83 in men and 0.81 in women. In hypopituitary patients with GH deficiency before pituitary replacement therapy, the mean utility was 0.67, and this improved by an average of 0.08 on treatment, the gain being independent of the number of deficient pituitary hormones in addition to GH. Although the baseline utility is lower than we have assumed (0.9), this has no effect on the ICER (£71 437 per QALY gained for baseline utility 0.8).

Key determinants of the cost-effectiveness include whether patients with secondary hypothyroidism go to the GP, whether the GP recognises patients with undetected hypopituitarism and hypothyroidism, the rate at which GPs request TFTs, the age of onset of hypopituitarism and the cost of an FT4 assay. Although the tornado graph suggests a modest effect of the cost of testing FT4, this is because our sensitivity analyses reflect only changes of 10%. Although the reagent cost of an FT4 test has fallen greatly in recent years, users still pay in the region of £3–£4 per test because of the additional costs noted above. We estimate that to make testing FT4 cost-effective at an ICER of £20 000 per QALY gained, the *total* laboratory cost to the user of measuring FT4 would need to be <£0.61.

Changing the probabilities of the GP recognising hypopituitarism, which in itself would entail costs, and the rate at which patients move from state 2 to state 3 have little or no effect on the ICER.

It may be argued that FT4 should be measured on only the first request for a particular patient, excluding requests for monitoring on patients with known thyroid disorders and reducing the proportion of patients who have a TFT request from 15.7% to 12%.⁶ However, these patients are also at risk of hypopituitarism, and so they probably should be included in the analysis.

One of the strengths of this study is estimating the probabilities of both false positives and false negatives. The proportion of false positives we chose at 0.33% is based on that found by Preiss *et al.*³ Our estimate of the proportion of false negative tests, that is, people in whom testing with normal TFTs despite hypopituitarism is based on Regal *et al.*²

One of the drawbacks of this study is that the modelled treatment effect depends on observational data. This is because few trials of treating hypopituitarism exist with

quality of life and mortality as endpoints. However, few clinicians would argue with the effectiveness of treating pituitary insufficiency. Moreover, our results are not sensitive to the magnitude of the treatment effect.

In conclusion, we have performed a cost-utility analysis of routinely measuring FT4 in addition to TSH in all specimens from primary care to identify patients with hypopituitarism. We find that the deterministic ICER for this strategy in someone aged 46 years compared with using TSH alone exceeding £71 000 per quality adjusted life year does not reflect good value given limited NHS resources. Adopting this would have a budget impact of about £19.1 million per year for England. Although this a small proportion of the annual budget of laboratories in England (estimated at around 4% of the total NHS budget, or about £2.5 billion),¹⁶ it represents money that could not be spent on other services that may have better returns for health. The magnitude of the change in QALYs between strategies is small, so the estimates are potentially unstable. Areas of uncertainty that markedly influence these estimates include the utility of treated and untreated hypopituitarism, the rate at which GPs request TFT in people without pituitary disease and the age of onset of hypopituitarism.

Contributors BS: conceived the research question, examined the existing literature, developed the software to analyse the options and produced graphical and numerical outputs, and wrote the initial drafts of the paper. AA and TJ: helped with the development of the research question and contributed to the writing and production of the paper.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Vanderpump MP, Tunbridge WM, French JM, *et al*. The incidence of thyroid disorders in the community: a twenty-year follow-up of the Wickham survey. *Clin Endocrinol* 1995;43:55–68.
- Regal M, Páramo C, Sierra SM, *et al*. Prevalence and incidence of hypopituitarism in an adult Caucasian population in northwestern Spain. *Clin Endocrinol* 2001;55:735–40.
- Preiss D, Todd L, Panarelli M. Diagnosing unsuspected hypopituitarism in adults from suggestive thyroid function test results. *Ann Clin Biochem* 2008;45:70–5.
- Jostel A, Ryder WDJ, Shalet SM. The use of thyroid function tests in the diagnosis of hypopituitarism: definition and evaluation of the TSH index. *Clin Endocrinol* 2009;71:529–34.
- Pappachan JM, Raskauskiene D, Kutty VR, *et al*. Excess mortality associated with hypopituitarism in adults: a meta-analysis of observational studies. *J Clin Endocrinol Metab* 2015;100:1405–11.
- Werhun A, Hamilton W. Thyroid function testing in primary care: overused and under-evidenced? A study examining which clinical features correspond to an abnormal thyroid function result. *Fam Pract* 2015;32:187–91.
- NICE, 2018. Available: <https://bnf.nice.org.uk/medicinal-forms/somatropin.html>
- Bolin K, Sandin R, Koltowska-Hägström M, *et al*. The cost-effectiveness of growth hormone replacement therapy (Genotropin®) in hypopituitary adults in Sweden. *Cost Eff Resour Alloc* 2013;11:24–4.
- Tomlinson JW, Holden N, Hills RK, *et al*. Association between premature mortality and hypopituitarism. West Midlands prospective hypopituitary Study Group. *Lancet* 2001;357:425–31.
- Office for National Statistics. The 21st century mortality files - deaths dataset, england and wales, 2017. Available: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/the21stcenturymortalityfilesdeathsdataset> [Accessed 2017].
- Koltowska-Hägström M, Kind P, Monson JP, *et al*. Growth hormone (GH) replacement in hypopituitary adults with GH deficiency evaluated by a utility-weighted quality of life index: a precursor to cost-utility analysis. *Clin Endocrinol* 2008;68:122–9.
- Williams C, Lewsey JD, Mackay DF, *et al*. Estimation of survival probabilities for use in cost-effectiveness analyses: a comparison of a multi-state modeling survival analysis approach with partitioned survival and Markov decision-analytic modeling. *Med Decis Making* 2017;37:427–39.
- Barth JH, Seth J, Howlett TA, *et al*. A survey of endocrine function testing by clinical biochemistry laboratories in the UK. *Ann Clin Biochem* 1995;32:442–9.
- Livingston M, Twomey PJ, Basu A, *et al*. Should free thyroxine go back into the routine thyroid profile? *Exp Clin Endocrinol Diabetes* 2015;123:594–7.
- Andela CD, Scharloo M, Pereira AM, *et al*. Quality of life (QOL) impairments in patients with a pituitary adenoma: a systematic review of QOL studies. *Pituitary* 2015;18:752–76.
- Lord Carter of Coles. Report of the review of NHS pathology services in England, 2006. Available: http://webarchive.nationalarchives.gov.uk/20061003103150/http://www.dh.gov.uk/PublicationsAndStatistics/Publications/PublicationsPolicyAndGuidance/PublicationsPolicyAndGuidanceArticle/fs/en?CONTENT_ID=4137606&chk=J4t5ZX